

# Statistical Properties of New Neutrality Tests Against Population Growth

Sebastian E. Ramos-Onsins and Julio Rozas

Departament de Genètica, Universitat de Barcelona, Barcelona

A number of statistical tests for detecting population growth are described. We compared the statistical power of these tests with that of others available in the literature. The tests evaluated fall into three categories: those tests based on the distribution of the mutation frequencies, on the haplotype distribution, and on the mismatch distribution. We found that, for an extensive variety of cases, the most powerful tests for detecting population growth are Fu's  $F_S$  test and the newly developed  $R_2$  test. The behavior of the  $R_2$  test is superior for small sample sizes, whereas  $F_S$  is better for large sample sizes. We also show that some popular statistics based on the mismatch distribution are very conservative.

## Introduction

Comparison of DNA sequences within and between species is a powerful approach not only for determining the evolutionary forces acting in specific gene regions but also for determining relevant aspects of the evolutionary history of the species (for reviews see, Takahata 1996; Rogers 1997; Harpending et al. 1998; Jorde, Bamshad, and Rogers 1998; Cann 2001). The coalescent theory (Kingman 1982a, 1982b; Hudson 1990; Donnelly and Tavaré 1995; Fu and Li 1999) is the most powerful theoretical approach for interpreting DNA sequence data. The coalescent is a population genetic model focused primarily on the neutral evolution of gene trees; this model provides the framework for the development of statistical tests and also provides very efficient computer simulations methods.

Tajima (1989b), Slatkin and Hudson (1991), and Rogers and Harpending (1992) pioneered the study of the effect of some demographic events on DNA sequence data. They have shown that a relatively recent demographic event, such as a population growth, causes most of the coalescent events to occur before the expansion and, consequently, samples of these populations have gene genealogies stretched near the external nodes and compressed near the root (i.e., star genealogies). Thus, population size changes can leave a particular footprint that may eventually be detected in DNA sequence data. This theoretical framework prompted the development of statistical tests for detecting population expansion.

The analysis of the distribution of pairwise differences, or mismatch distribution (Slatkin and Hudson 1991; Rogers and Harpending 1992), provides a method for inferring such demographic events. These authors have shown that, for nonrecombining DNA regions, constant size populations presented mismatch distributions with shapes with very little resemblance to that expected in growing populations. This prompted the development of some statistical tests for detecting expansion processes (Harpending et al. 1993; Harpending

1994; Eller and Harpending 1996; Rogers et al. 1996). One of the most frequently used tests is the raggedness statistic  $rg$  (Harpending et al. 1993). Although the distribution of the  $rg$  statistic is unknown, its confidence intervals could be obtained by computer simulations based on the coalescent algorithm. But because methods based on the mismatch distribution use little information accumulated in the data (Felsenstein 1992), tests based on the mismatch distribution should be very conservative.

In recent years, a number of authors have developed several methods of statistical inference and statistical tests using different approaches (e.g., Griffiths and Tavaré 1994; Bertorelle and Slatkin 1995; Rogers 1995; Aris-Brosou and Excoffier 1996; Fu 1996, 1997; Kuhner, Yamato, and Felsenstein 1998; Weiss and Von Haeseler 1998; Galtier, Depaulis, and Barton 2000; Furlong and Brookfield 2001). More recently, specific methods for detecting population expansions have also been developed for the analysis of microsatellite data (e.g., Kimmel et al. 1998; Reich and Goldstein 1998; Beaumont 1999; Reich, Feldman, and Goldstein 1999; King, Kimmel, and Chakraborty 2000).

Here we report the development of new statistical tests for detecting past population growth. We performed an extensive analysis of their statistical power against different alternative hypotheses, and we compared their relative performance with respect to others published in the literature. Although some authors (Braverman et al. 1995; Simonsen, Churchill, and Aquadro 1995; Fu 1996, 1997) have also investigated the power of some statistical tests against population growth and genetic hitchhiking (which leave similar footprints in DNA sequences), at present there is no exhaustive comparative analysis. The major population growth model investigated was the sudden (instantaneous) growth, although we also studied the power under the logistic model of population growth. The power of these tests was evaluated using random data sets generated by computer simulations based on the coalescent (Hudson 1990).

## Materials and Methods

We analyzed the performance of 17 statistical tests to distinguish specific models of population growth from the null hypothesis of a constant size population under

Key words: population growth, population expansion, coalescent simulations, neutrality tests.

Address for correspondence and reprints: Julio Rozas, Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, E-08071 Barcelona, Spain. E-mail: julio@bio.ub.es.

*Mol. Biol. Evol.* 19(12):2092–2100. 2002

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

the neutral model. Thus, we determined the power of these tests to reject the null hypothesis when the alternative hypothesis is really true. On the basis of the sequence information used, the test statistics evaluated have been classified into three major classes, namely classes I, II and III (see below). We developed several new statistical tests based on high-order moments (within classes I and III) because the distortion of the gene tree caused by the population growth would suggest that these types of tests could be more powerful than other tests available in the literature.

### Class I Statistics

Class I statistics use information of the mutation (segregating site) frequency. These statistics could be appropriate to distinguish population growth from constant size populations because the former generates an excess of mutations in external branches of the genealogy (i.e., recent mutations) and therefore an excess of singletons (substitutions present in only one sampled sequence) (Tajima 1989a, 1989b; Slatkin and Hudson 1991).

We studied the following test statistics: Tajima's  $D$ , and Fu and Li's  $D^*$ ,  $F^*$ ,  $D$  (named  $D^F$ ) and  $F$  statistics (Tajima 1989a; Fu and Li 1993; see also Simonsen, Churchill, and Aquadro 1995). These tests are based on the difference between two alternative estimates of the mutational parameter  $\theta = 2Nu$ , where  $N$  is the effective number of gene copies in the population (the number of females in the population for mtDNA regions or double the population size for an autosomal region) and  $u$  is the mutation rate. Tajima's  $D$  and Fu and Li's  $D^*$  and  $F^*$  statistics use information from only intraspecific data, whereas Fu and Li's  $D^F$  and  $F$  statistics use information from the number of recent mutations; the latter, therefore, requires the presence of an outgroup to be computed.

We developed a number of tests based on the difference between the number of singleton mutations and the average number of nucleotide differences. The  $R_2$  statistic is defined as

$$R_2 = \frac{\left(\sum_{i=1}^n \left(U_i - \frac{k}{2}\right)^2 / n\right)^{1/2}}{S} \quad (1)$$

where  $n$  is the sample size,  $S$  the total number of segregating sites,  $k$  the average number of nucleotide differences between two sequences, and  $U_i$  the number of singleton mutations in sequence  $i$ . The rationale of this test is that the expected numbers of singletons on a genealogy branch after a recent severe population growth event is  $k/2$ ; consequently, lower values of  $R_2$  are expected under this demographic scenario. The  $R_2$  statistic will be computed in the next version of the DnaSP (Rozas and Rozas 1999) software.

We also built two  $R_2$  related tests namely,  $R_3$  and  $R_4$ . These statistics differ from the  $R_2$  test in the power exponent values; in  $R_3$  and  $R_4$ , the exponent values of

2 and 1/2 (eq. 1) are replaced by 3 and 1/3, and by 4 and 1/4, respectively.

We have constructed three additional test statistics ( $R_{2E}$ ,  $R_{3E}$ , and  $R_{4E}$ ) that use information on the number of mutations in external branches; thus, an outgroup will be required for their estimation. The  $R_{2E}$  test is defined as

$$R_{2E} = \frac{\left(\sum_{i=1}^n \left(V_i - \frac{k}{2}\right)^2 / n\right)^{1/2}}{S} \quad (2)$$

where  $V_i$  is the number of external mutations in sequence  $i$ . The  $R_{3E}$  and  $R_{4E}$  tests differ from the  $R_{2E}$  test in the power values; the exponent values of 2 and 1/2 of equation (2) are replaced by 3 and 1/3, and by 4 and 1/4, respectively.

We have also developed two other tests ( $Ch$  and  $Che$ ) based on the difference between the number of singleton (and also for recent) mutations and their expected value:

$$Ch = \frac{(U - m)^2 S}{m(S - m)} \quad (3)$$

where

$$m = \frac{nk}{n - 1}$$

and  $U$  is the total number of singleton mutations. The  $Che$  test is constructed in the same way but by using information on the external mutations.

### Class II Statistics

In class II, we include statistical tests that use information from the haplotype distribution. We have only studied Fu's  $F_S$  test statistic (Fu 1997) within this class. This statistic, which is based on the Ewens' sampling distribution (Ewens 1972), has low values with the excess of singleton mutations caused by the expansion.

### Class III Statistics

Class III statistical tests use information from the distribution of the pairwise sequence differences (or mismatch distribution). It has been shown that population expansions leave a particular signature in the distribution of the pairwise sequence differences (Slatkin and Hudson 1991; Rogers and Harpending 1992); therefore, statistics based on the mismatch distribution can be used to test for demographic events. We evaluated the following statistics. (1) The raggedness  $rg$  statistic (Harpending et al. 1993; Harpending 1994). The raggedness statistic, which measures the smoothness of the mismatch distribution, differs among constant size and growing populations: lower  $rg$  values are expected under the population growth model. (2) The mean absolute error ( $MAE$ ) between the observed and the theoretical mismatch distribution (Rogers et al. 1996). (3) We also developed a new statistical test, the  $ku$  test, based on the fourth central moment (i.e., on the kurtosis) of the mismatch distribution. Given that population expansion

generates more smoothly peaked distributions, this statistic can distinguish between constant size and growing populations. Let  $d$ ,  $n_c$ , and  $W_i$  be the maximum number of differences in the mismatch distribution, the number of pairwise comparisons ( $= n(n-1)/2$ ), and the frequency of pairs of DNA sequences that differ by  $i$  mutations, respectively. We define:

$$ku = \frac{n_c(n_c + 1)q}{(n_c - 1)(n_c - 2)(n_c - 3)s^2} - \frac{3(n_c - 1)^2}{(n_c - 2)(n_c - 3)} \tag{4}$$

where

$$q = \sum_{i=0}^d W_i(i - k)^4, \text{ and}$$

$$s = \sum_{i=0}^d W_i(i - k)^2/(n_c - 1)$$

Empirical Distributions

We obtained the empirical distribution of each statistical test by Monte Carlo simulations based on the coalescent process for a neutral infinite-sites model, assuming a large population size (Kingman 1982a, 1982b; Hudson 1990). We also assumed that there is neither intragenic recombination nor migration and that the mutation rate is homogeneous across the DNA region. We performed the simulations conditional on the number of segregating sites ( $S$ ); that is, placing randomly  $S$  mutations along the tree (the so-called fixed  $S$  method). Given that the actual value of  $\theta$  is usually unknown, this method seems to be appropriate for testing purposes (Hudson 1993). The routine *ran1* (Press et al. 1992) was used as a random number generator. We conducted coalescent simulations for constant population size (null hypothesis) and for population growth (alternative hypothesis); the empirical distribution was estimated from 100,000 computer replicates for both the null and the alternative hypotheses.

For the constant size model (null hypothesis), the samples were generated using conventional procedures (Hudson 1990); in this model only two parameters are required: the sample size and the number of segregating sites. The sudden population growth model (Rogers and Harpending 1992) considers a population that was formerly at equilibrium, but  $t_e$  generations before the present one the population grew suddenly to the current size. Coalescent simulations under the sudden expansion model require four parameters:  $n$ ,  $S$ ,  $t_e$ , and  $D_e$ , where  $D_e$ , the degree of the expansion event, is:

$$D_e = N_{max}/N_0 \tag{5}$$

where  $N_{max}$  is the maximum population size (i.e., the current population size under the sudden expansion model) and  $N_0$  is the initial population size. For the simulations the  $t_e$  values were scaled in terms of  $N_{max}$  generations (denoted by  $T_e$ ). Coalescent simulations under

the sudden expansion model were performed by changing the time of the nodes as in

$$T_1 = \begin{cases} T, & T \leq T_e \\ T_e + \frac{T - T_e}{D_e}, & T > T_e \end{cases}$$

where  $T$  and  $T_1$  are the coalescence times (measured in  $N_{max}$  generations) under the constant size (i.e., the standard coalescent) and under the sudden expansion models, respectively (see Nordborg 2001). Under the later scenario, we generated samples using an extensive set of values of the parameter space.

We also conducted some coalescent simulations assuming that the population follows the logistic model of growth. In this model

$$N_T = N_0 + \frac{N_{max} - N_0}{1 + e^{-r(T_S - T - c)}} \tag{6}$$

where  $N_{max}$  is the maximum population size,  $N_T$  the population size at time  $T$  (the time is measured in  $N_{max}$  generations),  $r$  the growth rate,  $T_S$  the elapsed time (measured in  $N_{max}$  generations) from the beginning of the growth event, and  $c$  represents the reflection point of the growth curve (see eq. 16 in Fu 1997). It should be noted that under this model the current population size could be equal or lower than the  $N_{max}$ .

Coalescent simulations under the logistic model of growth were generated changing the times of the nodes according to the population size. These times are given by

$$T_1 = \frac{1}{N_c} \int_0^T N(x) dx \tag{7}$$

where  $T$  and  $T_1$  are the coalescence times (measured in  $N_{max}$  generations) of each node under the constant size and under the demographic models, respectively,  $N_c$  the current population size (i.e.,  $T = 0$ ) which can be obtained from equation 6, and  $N(x)$  the population size at time  $x$  (eq. 6). Therefore, we will compare two empirical distributions (under the null and the alternative hypotheses) with the same population size ( $N_c$ ) at the sampling time.

Critical Values and the Power of the Tests

We determined the critical values of each statistical test from its empirical distribution. The power of each test, or the probability of rejecting the null hypothesis (constant size population) when the alternative hypothesis (population growth) is true, was estimated as the proportion of computer replicates generated under the alternative hypothesis for which the null hypothesis was rejected. For the analysis, we fixed a significance level of  $\alpha = 0.05$ . Because the critical region for all alternative hypotheses would consist of only one side of the distribution, we conducted one-tailed tests. Specifically, all analyzed statistics, except *Ch*, *Che*, and *ku* had lower values under the population growth model.

Given that under the null hypothesis the empirical distribution of some statistics presented a reduced num-

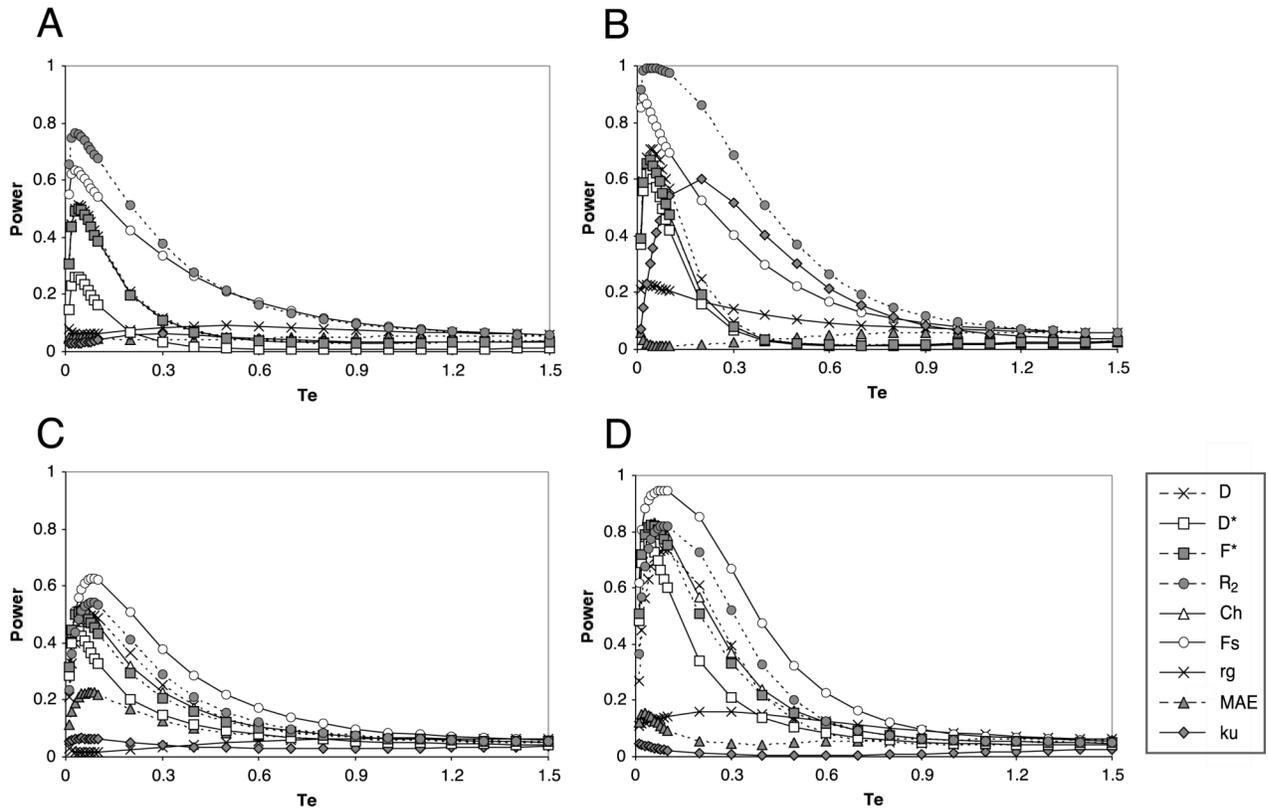


FIG. 1.—Effect of the elapsed time since the expansion event on the power of statistical tests. Results based on 100,000 sample replicates. (A)  $n = 10$ ,  $S = 10$ , and  $D_e = 100$ . (B)  $n = 10$ ,  $S = 50$ , and  $D_e = 100$ . (C)  $n = 50$ ,  $S = 10$ , and  $D_e = 10$ . (D)  $n = 50$ ,  $S = 50$ , and  $D_e = 10$ .

ber of points (e.g., the distribution of  $D^*$  statistic; see *Results*), the actual probability of rejecting the null hypothesis when it is true (i.e., the size of the test) could be lower than the nominal significance level of 0.05.

## Results

### Sudden Population Growth Model

We studied the power of 17 statistical tests under different values of  $n$ ,  $S$ ,  $D_e$ , and  $T_e$ . Although we have examined the power for a wide range of the parameter space, we will show only the most relevant cases (additional results and figures are available from the authors). The parameters fixed for illustrating the power were  $n = 10$  and  $n = 50$ ,  $S = 10$  and  $S = 50$ ,  $D_e = 10$  and  $D_e = 100$ , and  $T_e = 0.1$  (time for the maximum power; see below). These values give a clear view of the statistical power under some realistic cases: for small and big sample sizes, for a low and high number of mutations and for reasonable population growth parameters. In all cases, the parameter sets were chosen to avoid saturation of the power curves.

The power analysis of the tests  $R_3$ ,  $R_4$ ,  $R_{3E}$ , and  $R_{4E}$  show a similar power than the  $R_2$  and will not be presented here. Nevertheless, for some specific set of parameters the  $R_4$  and  $R_{4E}$  tests presented a slightly higher power than  $R_2$ . Generally, results of the statistical power of all statistical tests that use interspecific data presented a similar power than its equivalent statistic using intra-specific information (figures not shown).

Figure 1 shows the effect of  $T_e$ —the time elapsed since the expansion event—on the statistical power of different statistical tests. It can be observed that  $R_2$  and Fu's  $F_S$  are the most powerful tests: the  $R_2$  test is the most powerful for small sample sizes, whereas the behavior of Fu's  $F_S$  is better for large samples. The power of Tajima's  $D$  and Fu and Li's  $F^*$  is lower than  $R_2$  and  $F_S$ . The results also indicate that some commonly used tests based on the mismatch distribution,  $rg$  and  $MAE$ , are among the least powerful. All statistical tests show a peak in the statistical power at intermediate values of  $T_e$  ( $T_e \sim 0.1$ ); thus, it is unlikely to detect a population expansion when  $T_e$  is too small or too large. This result agrees with that obtained by Simonsen, Churchill, and Aquadro (1995) and Fu (1997).

The results of the effect of  $D_e$  on the power to reject the constant size model are depicted in figure 2. All statistical tests, except class III, increase the power to reject the constant size model with increasing  $D_e$ ; therefore, large samples will be needed to detect small population growth events. Again, tests based on the mismatch distribution are very insensitive in detecting population growths. The most powerful statistics are the  $R_2$  and the  $F_S$ . Tajima's  $D$ , Fu and Li's  $F^*$  and  $D^*$  and  $Ch$  have comparatively less power.

Figures 3 and 4 show the effect of the sample size and the number of segregating sites on the power to reject the neutral constant size model under specific alternative hypotheses. It should be expected that both

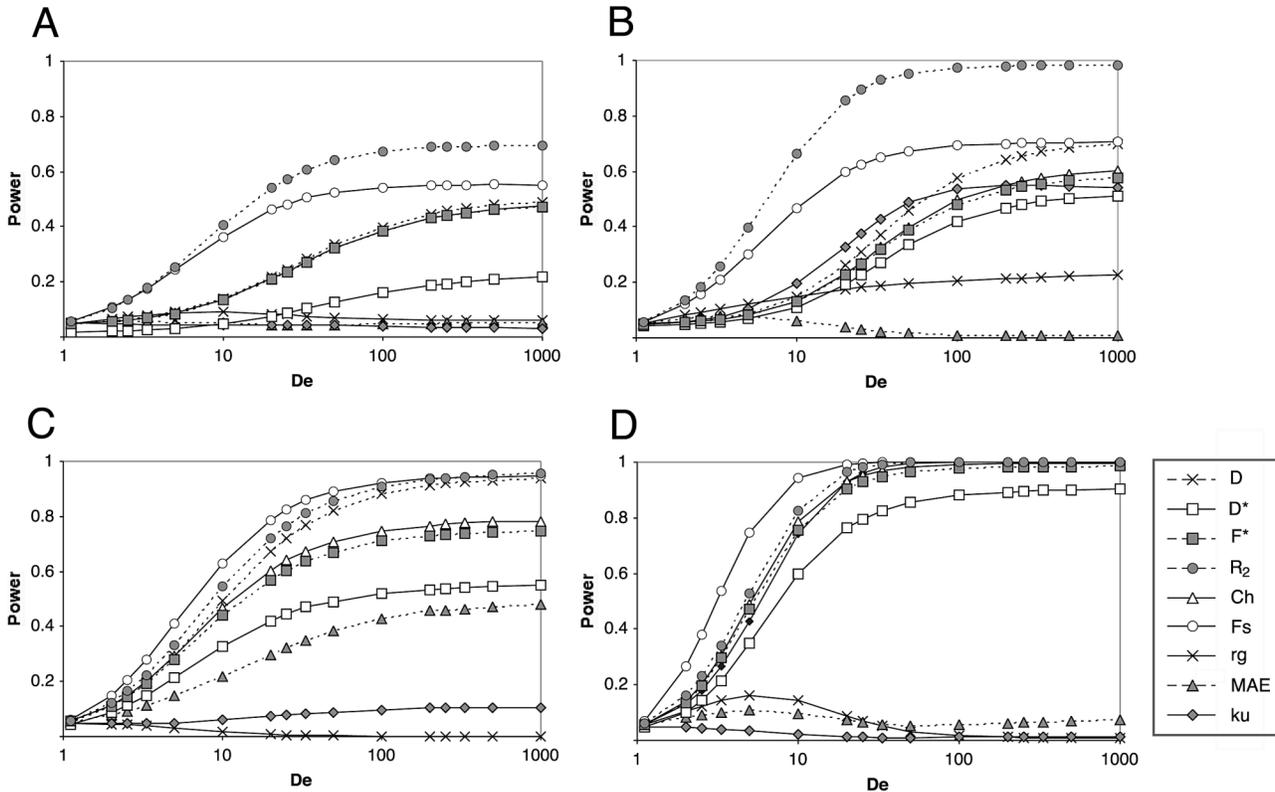


FIG. 2.—Effect of the degree of the expansion on the power of statistical tests. Results based on 100,000 sample replicates. (A)  $n = 10$ ,  $S = 10$ , and  $T_e = 0.1$ . (B)  $n = 10$ ,  $S = 50$ , and  $T_e = 0.1$ . (C)  $n = 50$ ,  $S = 10$ , and  $T_e = 0.1$ . (D)  $n = 50$ ,  $S = 50$ , and  $T_e = 0.1$ .

variables have a major effect on the statistical power, the larger the values of  $n$  or  $S$ , the more the power of the tests. But the effect on the power is different for different statistics: for small sample sizes (and a small number of segregating sites) the  $R_2$  statistical test is the most powerful (figs. 3A and 4A), whereas for larger sample sizes  $F_S$  is the most powerful one. Moreover, for small sample sizes the power of  $D^F$  and  $F$  is better than the counterpart tests without outgroup, although they are not as powerful as  $R_2$  and  $F_S$  (figure not shown). The results also indicate that statistical tests based on the mismatch distribution, the  $rg$ , and the  $MAE$  are among the least powerful. In fact, in some cases, the power decreases as the sample size increases.

It should be noted that statistics  $D^*$  (figs. 3A and 4B) and  $F_S$  (fig. 4A) have an irregular behavior because they show some atypical power drops with increasing sample size or the number of segregating sites. This unexpected pattern has two different explanations. In the case of Fu and Li's  $D^*$  statistic, the power drop is caused by a marked decrease in the actual significance level (i.e., the size of the test). In fact, the  $D^*$  empirical distribution has a reduced number of possible points causing, for some specific values, this level to drop to 0.02. The atypical pattern of the  $F_S$  test is due to the intrinsic structure of the statistic. In fact, the empirical distribution of  $F_S$  (both under  $H_0$  and  $H_1$  hypotheses) presents pronounced changes at specific ranges of values. That

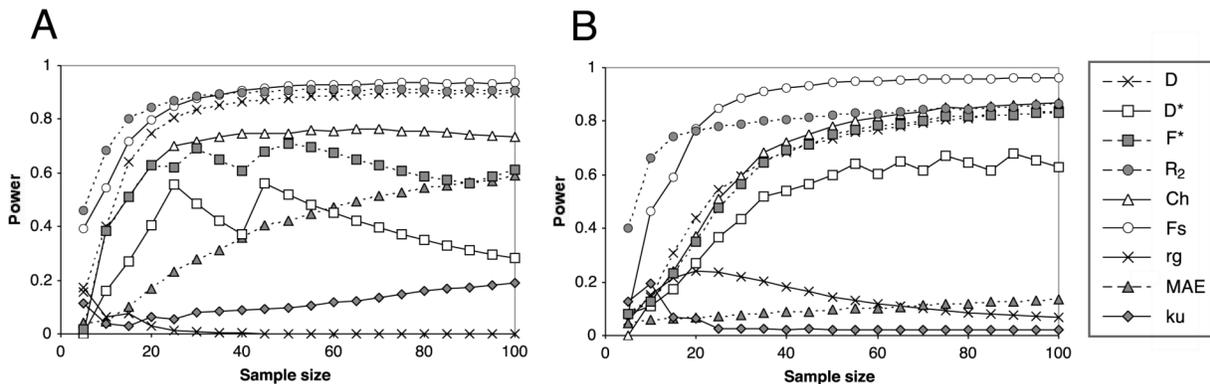


FIG. 3.—Effect of the sample size on the power of statistical tests. Results based on 100,000 sample replicates. (A)  $S = 10$ ,  $D_e = 100$ , and  $T_e = 0.1$ . (B)  $S = 50$ ,  $D_e = 10$ , and  $T_e = 0.1$ .

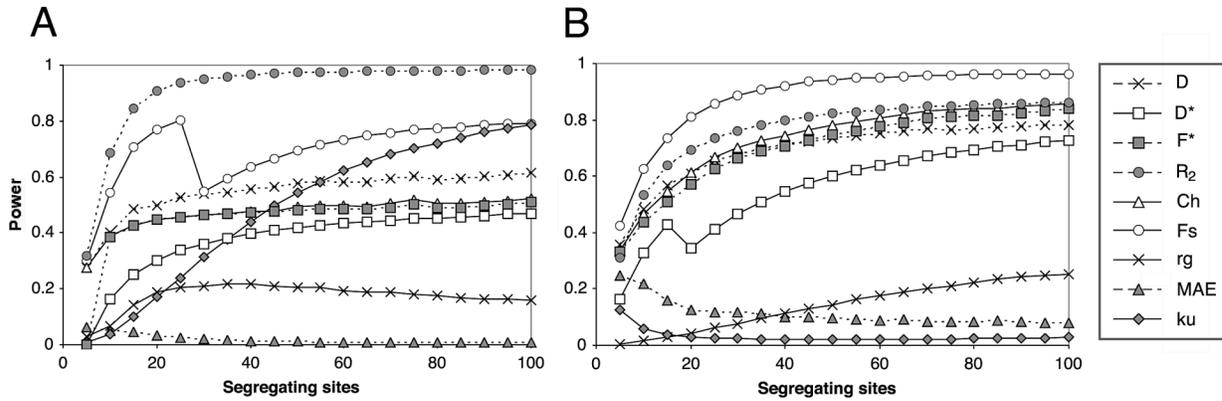


FIG. 4.—Effect of the number of segregating sites on the power of statistical tests. Results based on 100,000 sample replicates. (A)  $n = 10$ ,  $D_e = 100$ , and  $T_e = 0.1$ . (B)  $n = 50$ ,  $D_e = 10$ , and  $T_e = 0.1$ .

pattern causes marked changes in the power when these values are within the rejection region (results not shown). Nevertheless, this irregular behavior is not present in coalescent simulations conditional on the value of  $\theta$  (results not shown).

### Logistic Population Growth Model

We also conducted the analysis of power under a more realistic population growth scenario, the logistic population growth model. Using this model, we performed an explorative analysis of the most relevant cases to validate the conclusions of our work. We found that the assumption of the logistic population growth model does not change the major conclusions of the work. Even so, in comparison with the sudden growth model the maximum power of the tests is reached at higher values of the elapsed time; for instance, for the parameter sets used in Fu (1997) ( $r = 10$ ,  $c = 1$ ) the maximum power is at  $T_s \sim 1.2$ . In general, as expected (1) all statistical tests have less power under the logistic

than under the sudden growth models; nevertheless the decrease in the power is relatively uniform for all statistical tests and (2) the larger the value of  $r$ , the more power the tests have.

### Application to DNA Sequence Data

The present results have been applied to two published DNA data sets: the mtDNA variation analysis of a Turkish human population (Comas et al. 1996), and the survey of a human noncoding autosomal region (Alonso and Armour 2001). Comas et al. (1996) sequenced 360 base pairs of the region I of the mtDNA D-loop in 45 individuals. From the mismatch distribution analysis the authors suggested that the Turkish population had expanded recently. We determined the power of the different tests to identify which is most powerful against population growth. For the total data ( $n = 45$ ;  $S = 56$ ) and considering that  $D_e = 100$  and  $T_e = 0.4$  (scaled in terms of  $N$  generations) most tests were powerful enough, and several of them could reject the null hypothesis of constant size. We also determined whether the tests could also reject the null hypothesis for small sample sizes. For that, we reanalyzed a subset of 10 randomly chosen sequences from the data of Comas et al. (1996). Table 1 shows the estimates of the power and of  $P$  values of some statistical tests. The results clearly illustrate that the constant size hypothesis can be rejected by the most powerful tests (Fu's  $F_S$  and  $R_2$ ).

We also compared the  $P$  values and the power of the  $R_2$  and some of the statistical tests used in Alonso and Armour (2001). These authors performed a nucleotide variation study in 100 chromosomes sampled from different African and Euroasiatic populations. Although the surveyed region is autosomal, the Alonso and Armour (2001) results suggested that recombination should be reduced. We analyzed the Japanese population ( $n = 20$ ;  $S = 5$ ) using the same values of the recombination parameter  $R$  ( $R = 2N\rho$ , where  $\rho$  is the recombination rate per generation) as the published ones; for that analysis we used Hudson's (1983) algorithm to generate DNA samples under the coalescent with recombination (results based on 10,000 replicates). For the power analysis we consider that  $D_e = 100$  and  $T_e = 0.1$ . For  $R =$

**Table 1**  
Application of Some Statistical Tests to DNA Polymorphism Data

	$R_2$	$F_S$	$D$	$rg$
mtDNA Data (Comas et al. 1996)				
$n = 10$ ; $S = 21$				
Observed value . . . . .	0.0813	-5.585	-1.271	0.0523
$P$ value <sup>a</sup> . . . . .	0.009	0.003	0.107	0.243
Power <sup>b</sup> . . . . .	0.41	0.31	0.06	0.14
Autosomal data (Alonso and Armour 2001)				
$n = 20$ ; $S = 5$				
Observed value . . . . .	0.0923	-2.646	-1.140	0.147
$R = 0$ ; $P$ value <sup>a</sup> . . . . .	0.063	0.042	0.139	0.528
Power <sup>c</sup> . . . . .	0.48	0.57	0.46	0.01
$R = 1$ ; $P$ value <sup>a</sup> . . . . .	0.058	0.046	0.131	0.561
Power <sup>c</sup> . . . . .	0.50	0.58	0.47	0.01
$R = 10$ ; $P$ value <sup>a</sup> . . . . .	0.050	0.099	0.095	0.741
Power <sup>c</sup> . . . . .	0.53	0.42	0.67	0.0

<sup>a</sup> Probability of obtaining values equal or lower than the observed.  
<sup>b</sup> Power values assuming the sudden expansion model with  $\alpha = 0.05$ ,  $D_e = 100$ , and  $T_e = 0.4$ .  
<sup>c</sup> Power values assuming the sudden expansion model with  $\alpha = 0.05$ ,  $D_e = 100$ , and  $T_e = 0.1$ .

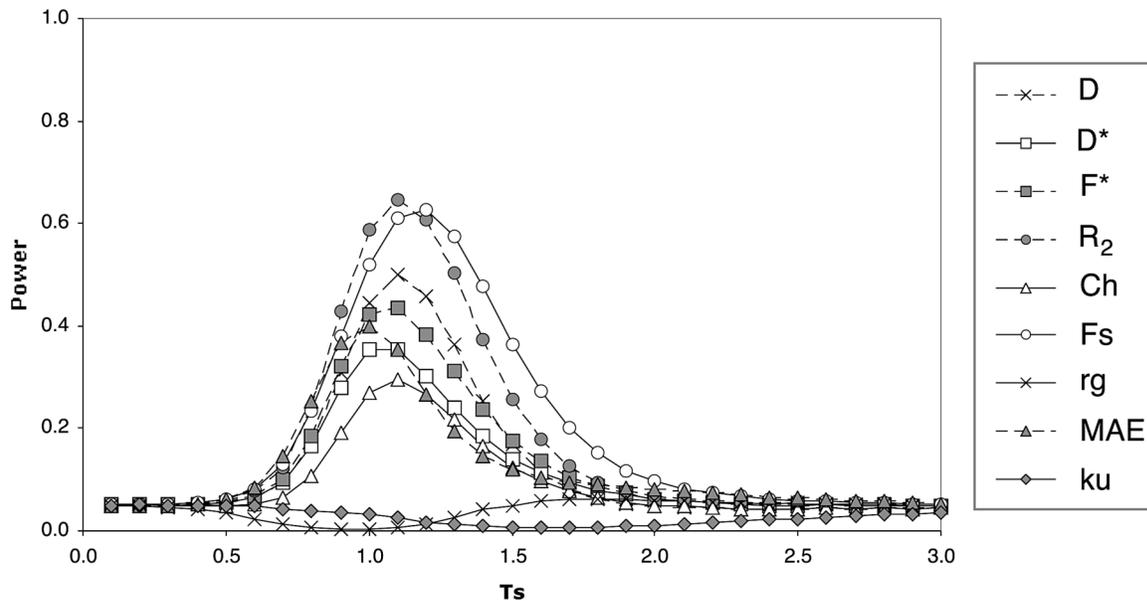


FIG. 5.—Effect of the elapsed time since the expansion event on the power of statistical tests under the logistic model of population growth. Results based on 100,000 sample replicates, fixing the value of  $\theta$  ( $\theta = 10$ ) with  $n = 50$ ,  $r = 10$ ,  $c = 1$ ,  $N_{max} = 20,000$  and  $N_0 = 1000$ .

0 (no recombination) only the  $F_S$  test can reject the null hypothesis of constant size. But for increasing recombination values the power of  $R_2$  and Tajima's  $D$  tests increases, whereas it decreases for  $F_S$  and  $rg$ . In fact, for  $R = 10$  only  $R_2$  allows the null hypothesis to be rejected.

## Discussion

In this article, we have examined the power of several statistical tests to determine which are most powerful in different population growth scenarios. The analysis has been performed by using a coalescent-based approach. There are other alternative approaches (likelihood-based methods) to study a population expansion process: the maximum likelihood (e.g., Griffiths and Tavaré 1994; Kuhner, Yamato, and Felsenstein 1998; Weiss and Von Haeseler 1998) and the Bayesian approaches (see Stephens 2001). The likelihood provides a framework for testing hypotheses; specifically, tests based on the likelihood ratio test statistic,  $\delta = -2 \ln(L_0/L_1)$ , where  $L_0$  and  $L_1$  are the maximum likelihood values under the null and the alternative hypothesis, can be used to discriminate between constant size and population growth. Unfortunately, the standard  $\chi^2$  approximation for the distribution of  $\delta$  might be inadequate. The empirical distribution of  $\delta$  could be generated, however, by computer simulation and from that distribution the critical values could also be obtained; nevertheless, this method is computationally very intensive.

We have shown that tests based on the mismatch distribution have little power against population growth. The  $MAE$  test is the less powerful one; although  $rg$  is more powerful than  $MAE$ , it works less well than nearly all class I and class II tests examined.  $ku$ , the newly developed test of class III, although better than  $MAE$

and  $rg$ , is clearly inferior to other class I and class II tests.

On the other hand, several class I and class II tests can detect population expansion even for small  $D_e$  values. We have shown that two of the surveyed tests ( $R_2$  and  $F_S$ ) are the most powerful for a variety of different conditions. These tests should therefore be chosen to test constant population size versus population growth. In particular, we suggest using the  $R_2$  statistical test for small sample sizes and  $F_S$  for large ones. Nevertheless, because  $R_2$  and  $F_S$  statistics use different kinds of information, discrepancies between these tests could provide information about the action of other evolutionary processes, for example on the intragenic recombination (see below).

Fu (1997) studied the power of some statistics under the logistic model of population growth. He conducted coalescent simulations fixing theta ( $\theta = 5$ ,  $\theta = 10$ ) instead of fixing  $S$ . To check the behavior of  $R_2$ , and other mismatch-based statistics, under these conditions we performed some additional simulations conditional on  $\theta$ . We found that the  $R_2$  and  $F_S$  are again the most powerful statistics (see an example in fig. 5). Interestingly,  $rg$  and  $MAE$  have better results fixing  $\theta$  than fixed  $S$ .

## Intragenic Recombination

The results from the present analysis are appropriate for nonrecombining DNA regions (i.e., mitochondrial or Y-chromosomal DNA regions). It is expected, however, that intragenic recombination substantially affects the power of the statistical tests surveyed (Rozas et al. 1999; Wall 1999). Indeed, a loss of power for those tests based on the haplotype distribution is expected (class II tests; e.g., Fu's  $F_S$  test) or for those based on the mismatch distribution (class III tests; e.g.,  $rg$  test).

The reason is that recombination, by shuffling nucleotide variation among DNA sequences (1) increases the number of haplotypes and (2) generates a much smoother mismatch distribution (Poisson-like). Consequently, class II and class III tests could be inadequate in detecting the signature left by a population growth on a recombining DNA region. Class I tests, on the contrary, should be less sensitive to intragenic recombination. To check our prediction, we conducted a few coalescent simulations using different values of the recombination parameter. Our preliminary results comparing the power of  $R_2$  and  $F_S$  tests show that the behavior of the former is better than the  $F_S$  for increasing levels of recombination (also see table 1).

#### Coalescent Simulations Conditional on the Number of Segregating Sites

The present power analyses have been performed conducting coalescent simulations conditional on the number of segregating sites. Given that the actual value of  $\theta$  is usually unknown, and that estimates of  $\theta$  are usually obtained from DNA polymorphism data information, the method seems to be appropriate (Hudson 1993; Depaulis, Mousset, and Veuille 2001; Wall and Hudson 2001). But Markovtsova, Marjoram, and Tavaré (2001) pointed out correctly that the power of coalescent-based tests are not independent of  $\theta$  and, therefore, the statistical power might vary as a function of  $\theta$  for a given  $n$  and  $S$ . To check that effect on the  $R_2$  we performed a prospective analysis generating samples conditional on  $\theta$  and  $S$  using the rejection algorithm of Tavaré et al. (1997). The results yield the same conclusions as that of Depaulis, Mousset, and Veuille (2001) and Wall and Hudson (2001), i.e., the fixed  $S$  method seems to be appropriate unless the actual value of  $\theta$  is far from Watterson's (1975) estimate of  $\theta$ .

#### Competitive Alternative Hypotheses

It should be stressed that a significant result (a significant departure from the null hypothesis) should be interpreted cautiously: there are several putative alternative hypotheses to single null hypotheses. Indeed, processes other than population expansion, such as genetic hitchhiking (Maynard Smith and Haigh 1974), could also produce similar genealogies (i.e., departures of the statistical tests in the same direction). Therefore, additional analyses could be necessary to discriminate between some competitive alternative hypotheses. For instance, because genetic hitchhiking in regions undergoing recombination will affect a relatively small fraction of the genome (close to the advantageous mutation), surveys at different gene regions across the genome could provide the opportunity to discriminate between population expansion and genetic hitchhiking (see Galtier, Depaulis, and Barton 2000).

To summarize,  $F_S$  and  $R_2$  are the best statistical tests for detecting population growth. The behavior of  $R_2$  is better for small sample sizes, whereas  $F_S$  is better for bigger sample sizes. Additionally, preliminary results also indicate that the behavior of  $R_2$  should be su-

perior when the intragenic recombination is considered. On the other hand, some popular statistics based on the mismatch distribution,  $rg$  and  $MAE$ , are very conservative.

#### Acknowledgments

We thank M. Aguadé, A. Navarro, H. Quesada, and C. Segarra for critical comments on the manuscript. This work was supported by grants PB97-0918 from the Dirección General de Investigación Científica y Técnica, Spain and 1999SGR-25 from the Comissió Interdepartamental de Recerca i Tecnologia, Catalonia, Spain, conferred on M. Aguadé.

#### LITERATURE CITED

- ALONSO, S., and J. A. L. ARMOUR. 2001. A highly variable segment of human subterminal 16p reveals a history of population growth for modern humans outside Africa. *Proc. Natl. Acad. Sci. USA* **98**:864–869.
- ARIS-BROSOU, S., and L. EXCOFFIER. 1996. The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Mol. Biol. Evol.* **13**:494–504.
- BEAUMONT, M. A. 1999. Detecting population expansion and decline using microsatellites. *Genetics* **153**:2013–2029.
- BERTORELLE, G., and M. SLATKIN. 1995. The number of segregating sites in expanding human populations, with implications for estimates of demographic parameters. *Mol. Biol. Evol.* **12**:887–892.
- BRAVERMAN, J. M., R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY, and W. STEPHAN. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphism. *Genetics* **140**:783–796.
- CANN, R. L. 2001. Genetic clues to dispersal in human populations: retracing the past from the present. *Science* **291**:1742–1748.
- COMAS, D., F. CALAFELL, E. MATEU, A. PÉREZ-LEZAUN, and J. BERTRANPETIT. 1996. Geographic variation in human mitochondrial DNA control region sequence: the population history of Turkey and its relationship to the European populations. *Mol. Biol. Evol.* **13**:1067–1077.
- DEPAULIS, F., S. MOUSSET, and M. VEUILLE. 2001. Haplotype tests using coalescent simulations conditional on the number of segregating sites. *Mol. Biol. Evol.* **18**:1136–1138.
- DONNELLY, P., and S. TAVARÉ. 1995. Coalescents and genealogical structure under neutrality. *Annu. Rev. Genet.* **29**:401–421.
- ELLER, E., and H. C. HARPENDING. 1996. Simulations show that neither population expansion nor population stationarity in a West African population can be rejected. *Mol. Biol. Evol.* **13**:1155–1157.
- EWENS, W. J. 1972. The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* **3**:87–112.
- FELSENSTEIN, J. 1992. Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. *Genet. Res.* **59**:139–147.
- FU, Y.-X. 1996. New statistical tests of neutrality for DNA samples from a population. *Genetics* **143**:557–570.
- . 1997. Statistical tests of neutrality against population growth, hitchhiking and background selection. *Genetics* **147**:915–925.
- FU, Y.-X., and W.-H. LI. 1993. Statistical tests of neutrality of mutations. *Genetics* **133**:693–709.

- . 1999. Coalescing into the 21st century:  $D_e$ , an overview and prospects of coalescent theory. *Theor. Pop. Biol.* **56**:1–10.
- FURLONG, R. F., and J. F. Y. BROOKFIELD. 2001. Inference of past population expansion from the timing of coalescence events in a gene genealogy. *J. Theor. Biol.* **209**:75–86.
- GALTIER, N., F. DEPAULIS, and N. H. BARTON. 2000. Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics* **155**:981–987.
- GRIFFITHS, R. C., and S. TAVARÉ. 1994. Sampling theory for neutral alleles in a varying environment. *Philos. Trans. R. Soc. Lond. B* **344**:403–410.
- HARPENDING, H. C. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol.* **66**:591–600.
- HARPENDING, H. C., M. A. BATZER, M. GURVEN, L. B. JORDE, A. R. ROGERS, and S. T. SHERRY. 1998. Genetic traces of ancient demography. *Proc. Natl. Acad. Sci. USA* **95**:1961–1967.
- HARPENDING, H. C., S. T. SHERRY, A. R. ROGERS, and M. STONEKING. 1993. Genetic structure of ancient human populations. *Curr. Anthr.* **34**:483–496.
- HUDSON, R. R. 1983. Properties of a neutral allele model with intragenic recombination. *Theor. Pop. Biol.* **23**:183–201.
- . 1990. Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* **9**:1–44.
- . 1993. The how and why of generating gene genealogies. Pp. 23–36 in N. TAKAHATA and A. G. CLARK, eds. *Mechanisms of molecular evolution*. Sinauer Associates, Inc., Sunderland, Mass.
- JORDE, L. B., M. BAMSHAD, and A. R. ROGERS. 1998. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *Bioessays* **20**:126–136.
- KIMMEL, M., R. CHAKRABORTY, J. P. KING, M. BAMSHAD, W. S. WATKINS, and L. B. JORDE. 1998. Signatures of population expansion in microsatellite repeat data. *Genetics* **148**:1921–1930.
- KING, J. P., M. KIMMEL, and R. CHAKRABORTY. 2000. A power analysis of microsatellite-based statistics for inferring past population growth. *Mol. Biol. Evol.* **17**:1859–1868.
- KINGMAN, J. F. C. 1982a. The coalescent. *Stoch. Proc. Applns.* **13**:235–248.
- . 1982b. On the genealogy of large populations. *J. Appl. Prob.* **19A**:27–43.
- KUHNER, M. K., J. YAMATO, and J. FELSENSTEIN. 1998. Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* **149**:429–434.
- MARKOVITSOVA, L., P. MARJORAM, and S. TAVARÉ. 2001. On a test of Depaulis and Veuille. *Mol. Biol. Evol.* **18**:1132–1133.
- MAYNARD SMITH, J., and J. HAIGH. 1974. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**:23–35.
- NORDBORG, M. 2001. Coalescent theory. Pp. 179–212 in D. J. BALDING, M. BISHOP, and C. CANNINGS, eds. *Handbook of statistical genetics*. John Wiley and Sons, West Sussex, U.K.
- PRESS, W. H., S. A. TEUKOLSKY, W. T. VETTERLING, and B. P. FLANNERY. 1992. *Numerical recipes in C. The art of scientific computing*. Cambridge University Press, New York.
- REICH, D. E., M. W. FELDMAN, and D. B. GOLDSTEIN. 1999. Statistical properties of two tests that use multilocus data sets to detect population expansions. *Mol. Biol. Evol.* **16**:453–466.
- REICH, D. E., and D. B. GOLDSTEIN. 1998. Genetic evidence for a paleolithic human population expansion in Africa. *Proc. Natl. Acad. Sci. USA* **95**:8119–8123.
- ROGERS, A. R. 1995. Genetic evidence for a pleistocene population explosion. *Evolution* **49**:608–615.
- . 1997. Population structure and modern human origins. Pp. 55–79 in P. DONNELLY and S. TAVARÉ, eds. *Progress in population genetics and human evolution*. Springer-Verlag, New York.
- ROGERS, A. R., A. E. FRALEY, M. J. BAMSHAD, W. S. WATKINS, and L. B. JORDE. 1996. Mitochondrial mismatch analysis is insensitive to the mutational process. *Mol. Biol. Evol.* **13**:895–902.
- ROGERS, A. R., and H. C. HARPENDING. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**:552–569.
- ROZAS, J., and R. ROZAS. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**:174–175.
- ROZAS, J., C. SEGARRA, G. RIBÓ, and M. AGUADÉ. 1999. Molecular population genetics of the *rp49* gene region in different chromosomal inversions of *Drosophila subobscura*. *Genetics* **151**:189–202.
- SIMONSEN, K. L., G. A. CHURCHILL, and C. F. AQUADRO. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**:413–429.
- SLATKIN, M., and R. R. HUDSON. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**:555–562.
- STEPHENS, M. 2001. Inference under the coalescent theory. Pp. 213–238 in D. J. BALDING, M. BISHOP, and C. CANNINGS, eds. *Handbook of statistical genetics*. John Wiley and Sons, West Sussex, U.K.
- TAJIMA, F. 1989a. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585–595.
- . 1989b. The effect of change in population size on DNA polymorphism. *Genetics* **123**:597–601.
- TAKAHATA, N. 1996. Neutral theory of molecular evolution. *Curr. Opin. Genet. Dev.* **6**:767–772.
- TAVARÉ, S., D. BALDING, R. C. GRIFFITHS, and P. DONNELLY. 1997. Inferring coalescence times for molecular sequence data. *Genetics* **145**:505–518.
- WALL, J. D. 1999. Recombination and the power of statistical tests of neutrality. *Genet. Res.* **74**:65–79.
- WALL, J. D., and R. R. HUDSON. 2001. Coalescent simulations and statistical tests of neutrality. *Mol. Biol. Evol.* **18**:1134–1135.
- WATTERSON, G. A. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**:256–276.
- WEISS, G., and A. VON HAESLER. 1998. Inference of population history using a likelihood approach. *Genetics* **149**:1539–1546.

WOLFGANG STEPHAN, reviewing editor

Accepted July 17, 2002